12-2019

The role of career and wage incentives in labor productivity: Evidence from a two-stage field experiment in Malawi

Hyuncheol KIM
Hong Kong University of Science and Technology

Seonghoon KIM
Singapore Management University, SEONGHOONKIM@smu.edu.sg

Thomas KIM
Ohio State University - Main Campus

Citation

KIM, Hyuncheol; KIM, Seonghoon; and KIM, Thomas. The role of career and wage incentives in labor productivity: Evidence from a two-stage field experiment in Malawi. (2019). *Review of Economics and Statistics*. 102, (5), 839-851.

Available at: https://ink.library.smu.edu.sg/soe_research/2320

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

The Role of Career and Wage Incentives in Labor Productivity: Evidence from a Two-stage Field Experiment in Malawi[1]

Hyuncheol Bryant Kim[2], Seonghoon Kim[3], and Thomas T. Kim[4]

15 June 2019

Abstract

We study how career and wage incentives affect labor productivity through self-selection and incentive effect channels using a two-stage field experiment in Malawi. First, recent secondary school graduates were hired with either career or wage incentives. After employment, a half of

[1] We are grateful to the following staff members of Africa Future Foundation for their field assistance: Narshil Choi, Jungeun Kim, Seungchul Lee, Hanyoun So, and Gi Sun Yang. We also thank Derek Lougee and Seollee Park for excellent research assistance. In addition, we thank Syngjoo Choi, Andrew Foster, Dan Hamermesh, Guojun He, Kohei Kawaguchi, Asim Khwaja, Etienne Lalé, Kevin Lang, Suejin Lee, Pauline Leung, Zhuan Pei, Cristian Pop-Eleches, Nick Sanders, Slesh Shrestha, and Jungmin Lee as well as seminar participants at various seminars and conferences. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant. All errors are our own.

[2] hk788@cornell.edu; Department of Policy Analysis and Management, Cornell University

[3] seonghoonkim@smu.edu.sg; School of Economics, Singapore Management University

[4] kim.6903@osu.edu; Department of Agricultural, Environmental, and Development Economics, Ohio State University

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

workers with career incentives randomly received wage incentives, and a half of workers with wage incentives randomly received career incentives. Career incentives attract higher-performing workers than wage incentives, but do not increase productivity conditional on selection. Wage incentives increase productivity for those recruited through career incentives. Observable characteristics are limited in explaining selection effects of entry-level workers.

Keywords: Career Incentive, Wage Incentive, Internship, Self-selection, Labor Productivity

JEL Classification: J30, O15, M52

1. Introduction

Work incentives are essential tools to improve labor productivity. Firms try to recruit productive workers and motivate existing employees to exert more effort through work incentives. Career incentives (tenure and promotion) and financial incentives (higher wage, cash bonus, and employee stock option) are common examples of work incentives. There are two channels through which work incentives can affect labor productivity: *selection* and *incentive effects*.[5] A better understanding of how different incentives affect labor productivity would enable firms to design optimal hiring and compensation strategies that maximize labor productivity and reduce the need for costly screening processes.

We provide experimental evidence on how career and wage incentives affect labor productivity through self-selection and incentive effect channels. We conduct a two-stage

[5] The *incentive* effect refers to the difference in labor productivity when incentives affect performance holding employee composition constant. The *selection* effect refers to the difference in labor productivity driven by workers' self-selection into the job.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

randomized controlled trial to separately isolate the selection and incentive effects of these incentives in collaboration with Africa Future Foundation (AFF), an international non-governmental organization (NGO), in the context of a recruitment drive for entry-level enumerators for a population census survey in rural Malawi.

The career incentives we study consist of a future job prospect and a recommendation letter, which are typical benefits of an internship position.[6] The wage incentives in our study are composed of a lump-sum salary and performance-related bonus payment. Firms might expect that career incentives attract workers more forward-looking and/or risk-loving than others because an internship position implies taking the risk of not being employed at the end of the internship. On the other hand, firms might expect that wage incentives attract workers more extrinsically motivated by monetary compensation.

Our research setting, the recruitment of entry-level enumerators in Malawi, is suitable to study the role of work incentives in productivity because we are able to measure high frequency individual-level labor productivity. The nature of an enumerator job is multidimensional because enumerators are expected to conduct interviews both quickly and accurately. Thus, we measure job performance by the number of surveys conducted per day (survey quantity) and the proportion of errors/mistakes made in a survey (survey quality). In addition, our setting has advantages to study the role of work incentives especially in worker self-selection. Worker screening in

[6] An internship is a temporary position that can be paid or unpaid, and is distinguished from a short-term job in that it emphasizes on-the-job training for students or entry-level workers. Internship programs are widely available in Malawi in the public, private, and NGO sectors. For example, about 20% of regular workers in AFF are hired through the internship program.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

developing countries is difficult because observable information on worker skills such as certification, accreditation, and the past work history are limited. It is even more challenging to observe the productivity of entry-level workers due to no or short work history.

To hire enumerators, AFF approached 440 randomly selected recent high school graduates in its project areas. As shown in Figure 1, in the first stage, study subjects were randomly assigned to one of the two groups: (i) those who received a job offer with career incentives (hereafter the *Internship* group) and (ii) those who received a job offer with wage incentives (hereafter the *Wage* group). Those assigned to the *Internship* group received an internship opportunity that comes with (a) a potential long-term employment opportunity at AFF as a regular employee and (b) a recommendation letter specifying their job performance.[7] A one-time temporary work opportunity with a lump-sum wage and a bonus payment based on job performance was offered to those assigned to the *Wage* group.

Individuals who accepted the job opportunity in the first stage proceeded to enumerator training and the second-stage randomization. After completing the training, a randomly selected half of the job takers in the *Internship* group additionally received the same wage incentives of the *Wage* group without prior notice. In the same manner, a randomly selected half of the job takers in the *Wage* group additionally received the same career incentives of the *Internship* group without

[7] An entry-level regular position (enumerator or data entry clerk) at AFF has career advancement prospects that lead to more advanced positions. AFF did not explicitly state the actual probability of being hired to the *Internship* group. We acknowledge that changing probabilities of being hired after the internship might affect effort levels, but we do not compare different levels of the same incentive, but rather two different types of incentives.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

prior notice. As a result, this research design creates four sub-groups: *Group 1 (G1)* and *Group 2 (G2)* became enumerators through career incentives, but only *G2* received additional wage incentives. Similarly, *Group 3 (G3)* and *Group 4 (G4)* became enumerators through wage incentives, but only *G3* received additional career incentives.

We isolate the selection effect on labor productivity by comparing *G2* and *G3*, both of which have identical incentives (both career and wage incentives) during the work period. However, the channels through which they were attracted to the job are different.[8] Our identifying assumption of the selection effect is that sequences in which first-stage and second-stage incentives are presented to G2 and G3 participants are independent of the combined value of the career and wage incentives. This assumption is required both in the conceptual framework (Conceptual Framework Appendix A.2) and the empirical analysis (Section 4.3). We discuss the reliability of this assumption with further details in Section 4.3.

In addition, we estimate the incentive effects of wage incentives (henceforth, wage incentive effects) on job performance among the job takers in the *Internship* group by comparing *G1* and *G2*. Both groups became enumerators through the career incentives, but only *G2* received additional wage incentives. Hence, any difference in performance between *G1* and *G2* can be interpreted as wage incentive effects among the job takers in the *Internship* group. Similarly, we estimate the incentive effects of career incentives (henceforth, career incentive effects) on job performance among the job takers in the *Wage* group by comparing *G3* and *G4.* Any difference in

[8] The comparison of *G2* and *G3* can be also interpreted as the selection effect of the wage incentives evaluated against the career incentives, but for the sake of convenience, we focus on the career incentives.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

performance between *G3* and *G4* can be interpreted as career incentive effects among the job takers in the *Wage* group.

Of 440 randomly selected recent male high school graduates whom AFF approached for the baseline survey of this study without prior notice of job opportunity, 362 (82.3%) participated in the baseline survey.[9] Of 176 study participants assigned to the *Wage* group, 74 (42.0%) accepted a job offer by joining the training session. Of 186 study participants assigned to the *Internship* group, 74 (39.8%) took up the job offer. Of 148 trainees, 11 dropped out from the training. As a result, 137 enumerators worked in the field for an average of 18 days interviewing 21,561 households.[10]

We reach four main conclusions using data on labor productivity measured by survey quality and survey quantity. First, we find that career incentives, compared to wage incentives, attract workers with higher labor productivity through the self-selection mechanism. Second, we find that the incentive effects of career incentives among those recruited by wage incentives are

[9] There were 536 eligible study subjects who were male and recent high school graduates in AFF's project areas. Of the 536, AFF provided job offers to a randomly selected group of 440. The other 96 subjects were also invited to participate in the baseline survey, although they did not receive a job offer. Individual characteristics and the balance between the two groups (440 vs. 96) are shown in Table A.1.

[10] Throughout this paper, *target study participants* refer to the 440 individuals who were invited to participate in the baseline survey, *study participants* refer to the 362 individuals who participated in the baseline survey; *trainees (job takers)* refer to the 148 individuals who joined the training; and *enumerators* refer to the 137 individuals who worked in the field.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

limited in improving labor productivity. Third, we find that wage incentives causally increase labor productivity among those recruited through career incentives. As a result, overall job performance is highest among *G2* enumerators who were hired through the career incentive channel and additionally received wage incentives. Lastly, we find that observable individual characteristics are limited in explaining the selection effect of entry-level workers, suggesting a limitation of screening based on observable characteristics and a need for a self-selection mechanism that can attract productive workers with desirable (unobserved) characteristics.

Our primary contribution to the literature is that we study career and wage incentives, the most common types of work incentives, jointly in the same setting, and provide real-world evidence on how these incentives affect labor productivity by identifying the selection and incentive effect channels through two-stage randomization.

Previous studies estimating the selection and incentive effects separately focus only on financial incentives (Lazear, 2000; Gagliarducci and Nannicini, 2013; Guiteras and Jack, 2018). Moreover, their findings on relative importance of selection and incentive effects are mixed. For example, Lazear (2000) isolates worker selection and incentive effects of pay-for-performance using non-experimental panel data on job performance from a large manufacturing factory in the US. He shows evidence that the change to piece rate pay increases labor productivity by 44% with a half of it coming from the selection effect and the other half comes from the incentive effect. Gagliarducci and Nannicini (2013) also identify the selection and incentive effects of wage incentives on the performance of politicians by exploiting policies that discontinuously change their salaries and limit political terms. They find that a higher wage attracts more educated candidates and leads to improved efficiency of public finance through the selection channel. By contrast, Guiteras and Jack (2018) find evidence from bean-sorting workers in rural Malawi that a

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

higher piece rate increases productivity only through the incentive effect channel, not through the worker selection channel. Our results are consistent with Lazear's (2000) findings that both selection and incentive effects are important.

There are several studies focusing on the selection effects of work incentives. Dohmen and Falk (2011) show that sorting of workers largely explains higher labor productivity under a variable-payment scheme compared to a fixed-payment scheme in a laboratory experiment setting. Dal Bó et al. (2013) show that a higher wage attracts more qualified applicants without the cost of losing workers with strong public service motivation in a recruitment drive for Mexico's public sector workers. Ashraf et al. (2016) similarly show that salient career incentives attract more productive workers without discouraging those with pro-social preferences from applying for a job in a recruitment drive for community health workers in Zambia. On the other hand, Deserranno (2018) finds that the expectation of a higher salary for a newly created health-promoter position discourages job applications from socially motivated candidates in Uganda. While the previous literature estimated selection effects of either financial incentives or career incentives, we estimate selection effects of career incentives evaluated against wage incentives.

In addition, our study is related to another strand of the literature on incentive effects on job performance. The previous literature mainly focuses on financial incentives, to the best of our knowledge (Gneezy and List, 2006; Shearer, 2004; Glewwe et al., 2010; Duflo et al., 2012; Fryer, 2013; Ashraf et al., 2014). For example, Gneezy and List (2006) empirically test the gift exchange theory developed by Akerlof (1984) and show that workers exert more efforts when they receive a financial incentive ("gift") from their employers. Shearer (2004) presents experimental evidence from Canadian tree planters that piece rates induce more effort than do fixed wages. By contrast, ours is the first of its kind to estimate career incentive effects.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Lastly, our study is related to the literature on internships. Most existing studies on internships are descriptive (Brooks et al., 1995; D'Abate et al., 2009; Liu et al., 2014). A rare exception is Nunley et al. (2016), which sends out fake résumés with randomly changed characteristics of applicants. They find that a résumé with internship experience receives 14% more callbacks from potential employers. However, a major limitation of the résumé audit study is lack of job performance data. Since career incentives in this study closely follow the structure of an (unpaid) internship program in the real world, this study offers experimental evidence on the effects of an internship on worker selection and job performance.

The remainder of the paper is structured as follows: Section 2 outlines the research context and design. Section 3 describes the data and reports sample statistics. Section 4 presents the main results on labor productivity and discusses the findings. Section 5 concludes.

2. Research Context and Design

2.1. Research Context

Malawi is one of the least developed countries in the world with GDP per capita in 2015 of US$382 (World Bank, 2016). Among 20–29 years old males, 19.6% completed secondary school education according to the 2010 Malawi Demographic and Health Survey. Employment in the official sector is 11% and the median monthly income is US$28.8 (13,420 MWK) (National Statistical Office of Malawi, 2014).[11]

AFF conducted a district-wide population census of Chimutu, a rural district located

[11] MWK denotes Malawi Kwacha. As of January 1, 2015, US$1 was equivalent to 466 MWK. Throughout the paper, we use this as the currency exchange rate.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

outside of the capital city of Malawi, in January 2015. Chimutu district consists of 52 catchment areas with about 94,000 people (around 24,000 households). AFF planned to complete a census within a month by hiring more than 130 enumerators.

The enumerator position could be an attractive starting job for entry-level young workers because it offers a competitive salary and confers career-advancing incentives. For example, AFF's many regular staff members were initially recruited as enumerators. The role of the census enumerators was to interview household heads to collect basic demographic, socioeconomic, and health information. During the census period, enumerators stayed at a house in the assigned catchment area rented by AFF. Since enumerators interviewed many residents in remote villages to collect a variety of personal and complex information, the job required both cognitive and interpersonal skills as well as physical endurance.

Study participants to whom AFF offered the enumerator job were drawn from the sample of individuals who participated in the 2011 secondary school student survey in four districts in Malawi, including Chimutu. This 2011 survey was a baseline survey for AFF's previous research program that randomly provided HIV/AIDS education, male circumcision, and financial support for female education in their catchment areas.[12] Of the 536 males who participated in the 2011 secondary school survey and graduated from secondary school in July 2014, AFF randomly selected 440 as target study participants. 362 study participants participated in the survey (i.e., the baseline survey of this study) without prior notice of a potential job offer. This sample recruitment approach allowed AFF to hire workers familiar with the census area. AFF considered only males

[12] AFF's catchment areas include the following four districts: Chimutu, Chitukula, Tsbango, and Kalumba. For details of AFF programs, see Data Appendix A.4.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

due to security concerns in the field. In addition, AFF required secondary school graduation as proof of minimum cognitive skill requirements.

Outside options for the enumerator job are other formal sector jobs, household farming, and repeating secondary school. For instance, at the time of the baseline survey, 4.7% of our study participants were working for pay in formal sectors, 4.3% were working for their family business (mainly farming), and 15.8% were attending vocational schools or colleges. About 60% were actively searching for jobs.

Our sample recruitment strategy has two advantages. First, we observe the population of a young cohort whose members are potentially interested in a job opportunity in the local labor market, contrary to existing studies that observe only job applicants. This feature of our sampling allows our findings to have greater external validity by addressing the concern that individual characteristics of job applicants may be systematically different from those of non-applicants. For example, applicants could be more likely to possess the necessary skills, have better access to the information (at least for a job vacancy), and/or be less likely to be happy with their existing positions if they are currently working for another employer. Hence, the estimation of selection effects of any work incentives is inherently local to job applicants. Second, approaching those who just graduated from secondary school is relevant to an internship, which mainly targets young and entry-level workers.

2.2. Experimental Design

In this section, we explain the details of the experiment. The discussion of a conceptual framework that motivates our experimental design and provides guidelines for the empirical analysis is in the Appendix.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

2.2.1. Baseline survey and first-stage randomization

We describe the research stages in chronological order as shown in Table 1. As stated in the introduction, AFF invited 440 males who met the eligibility criteria (target study participants) for the baseline survey (Row A) and 362 (82.3%) participated in the baseline survey (Row B).[13] In addition, AFF invited study participants soon after the census was completed between April and June 2015 to measure time and risk preferences and rational decision-making ability.[14]

To minimize unexpected interaction among workers with different incentives, first-stage randomization was performed in advance, and the baseline survey and training were also conducted separately for the *Internship* group and the *Wage* group. Study participants were given a job offer with detailed information on an enumerator position at the end of the baseline survey. It is noteworthy that a job offer was valid conditional on successful completion of the training. We refer to a conditional job offer simply as a job offer henceforth. Study participants were not aware of the other type of incentives when they received an offer.

Of 220 target study participants assigned to the *Wage* group, 176 (80.0%) showed up for the baseline survey (Row B) and were given a short-term (verbal) job offer, each with a fixed salary of 10,000 MWK (US$21.5) for up to 30 days and performance pay of 500 MWK (US$1.1)

---

[13] Those who did not participate in the survey were unreachable (45%), refused to participate (13%), or could not participate in the survey because they were at school (32%) or working (10%).

[14] This survey was conducted to measure time and risk preferences and rational decision-making ability after the census was completed under the assumption that these measures are not affected by our interventions. Out of 440 target study participants, 334 (76%) participated in the survey. We further discuss the data collected from these surveys in Section 3.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

for every extra 8 households after the first 160 households.[15] Of 220 target study participants assigned to the *Internship* group, 186 (84.5%) showed up for the baseline survey (Row B) and were given a (verbal) job offer with career incentives which consist of a recommendation letter and the prospect of working at AFF as a regular staff member.

The base wage of 10,000 MWK (US$21.5) was competitive for young workers who had just graduated from secondary schools because the median monthly salary of secondary school graduates in 2013 was 12,000 MWK (US$25.8), according to the Malawi Labor Force Survey (NSO, 2014).[16] AFF notified the *Internship* group that there would be a chance of a long-term contract, without specifying the precise probability, depending on job performance during the contract period and AFF's job vacancies. Working as an intern without knowing the exact probability of hiring is close to the general internship setting. Lastly, one-time transportation support, on average about 1,500 MWK (US$3.2), was given to both *Wage* and *Internship* groups depending on the distance from the worker's home and the dispatched village.

2.2.2. Training

Those who took the job offer were required to participate in a 1-week training program in

[15] This rule gives an impression to enumerators that surveying 160 households is the *de facto* expectation of good performance. We acknowledge that this reference could increase or decrease average survey completion, but having a specific rule or a cut-off point about performance is unavoidable if an organization has to offer rule-based performance pay.

[16] The prospect of a regular entry-level staff position at AFF whose entry-level monthly salary is 26,000 MWK (US$55.8) could be attractive.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

January 2015. It was designed to equip trainees with the necessary skills and knowledge for the census work. The training outcomes were measured by a quiz score and the proportion of erroneous entries in a practice survey. To prevent interaction between participants with different incentives, the *Internship* group (the first week) and *Wage* group (the second week) joined the training sessions separately, but the instructors and the training materials were identical.

Out of the 186 study participants in the *Internship* group, 74 (39.8%) participated in the training session, as did 74 out of 176 (42%) study participants in the *Wage* group (Row C). The job take-up rates (training participation rates) between the *Internship* group and the *Wage* group were not statistically different. However, 11 trainees from the *Internship* group were not hired because of low training performance, while no one failed from the *Wage* group (Row D). In total, 137 enumerators were finally hired, 63 of which were from the *Internship* group and 74 from the *Wage* group (Row E). As a result, we do not observe job performance of 11 trainees from the *Internship* group who failed the training requirement.[17]

2.2.3. Second-stage randomization

Second-stage randomization was conducted during the training, and the randomization results were announced after the training completion but before the dispatch to the catchment area. The wage incentives were given to a randomly selected half of the *Internship* group. Similarly, the career incentives were given to a randomly selected half of the *Wage* group. The second-stage randomization was announced publicly. Therefore, both G1 and G2 enumerators learned about the additional wage incentives, and both G3 and G4 enumerators learned about the additional career incentives. AFF staff explained to enumerators that they would distribute additional incentives in

[17] We discuss this further in footnote 36.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

a random manner due to budget constraints. No enumerators refused to accept the additional incentives, which implies that the composition of worker characteristics between *G1* and *G2* and between *G3* and *G4* remains the same.

Right after the second-stage randomization, AFF supervisors and enumerators had a one-on-one session to explain the details of the contract, and the enumerators signed the employment contract as shown in Figures A.1, A.2, and A.3.[18] To illustrate, the employment contract of *G1* explicitly states that enumerators will not be given any financial compensation and will be provided with a recommendation letter and a potential job opportunity based on their performance.

2.2.4 Census and post-enumeration survey

Enumerators were dispatched to 52 catchment areas in January 2015. They were randomly assigned to catchment areas stratified by population and land size, and worked independently. Enumerators in the same catchment area have the same incentives to prevent unexpected peer effects. In addition, enumerators were not assigned to areas from which they originally came, as locality could affect their performance. The census survey took about 25 minutes on average to interview a household head. Enumerators were expected to survey at least eight households per day. In total, enumerators surveyed 21,561 households during the contract period.

AFF supervisor teams, which consisted of two supervisors per team, visited enumerators to monitor and guide enumeration work on randomly selected dates without prior notice. Supervisors are AFF's regular staff members, each with at least 3 years of experience conducting field surveys. AFF randomly assigned five supervisor teams to 52 catchment areas for their visits.

---

[18] Through the one-on-one meeting, AFF explained to G4 enumerators that their position would be a one-time employment opportunity even though it was not explicitly mentioned in the contract.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Most enumerators met a supervisor team at least once during the census period; 37% of the enumerators met supervisors twice and the remaining 60% met supervisors once. Enumerators were aware of supervisor visits but did not know the exact date. Supervisors joined each enumerator for interviews of about three households, addressed common errors, and provided overall comments at the end of the visit.

Shortly after the completion of the census, AFF conducted a post enumeration survey (PES) to correct errors found in the original census interview, find omitted households, and measure subjective performance evaluation (SPE) by revisiting all households in Chimutu. AFF announced a PES plan to evaluate the performance before the field dispatch to prevent enumerators from outright cheating or fabricating census interview sheets.[19]

As stated in the employment contract, AFF provided recommendation letters to the enumerators with career incentives (*G1*, *G2*, and *G3*) in May 2015. The recommendation letter was signed jointly by the director of AFF and the head of the Chimutu district. The letter specified the job description of an enumerator and his relative job performance.[20]

[19] Hiring enumerators as regular staff members required the calculation of job performance after the completion of the census, which can take at least two months. Meanwhile, AFF hired 43 PES enumerators among 98 census enumerators with career incentives (*G1*, *G2*, and *G3*) on a temporary basis (2–3 months) through a simple performance evaluation based on SPE by supervisors and error rates measured from five randomly selected surveys.

[20] If an enumerator has higher job performance than the average, the letter specifies a very strong recommendation. If an enumerator has performance below the average, the letter specifies a somewhat lukewarm recommendation.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

3. Data

We use data from various sources, including baseline and follow-up surveys, administrative data on training and job performance, and the Chimutu population census. First, we use data from the 2011 secondary school student survey. It contains rich information on a variety of areas covering demographics, socioeconomic status, health, and cognitive ability. Second, we use data from the 2014 baseline survey, which collects information on demographics, education, employment history, cognitive abilities, non-cognitive traits, and HIV/AIDS related outcomes.

We measure cognitive ability in two distinct ways. The first measure is Math and English scores of the 2014 Malawi School Certificate of Education (MSCE) test, which are easily observable in the local labor market.[21] The second measure is the scores of Raven's matrices test and the verbal and clerical ability tests of the O*NET, which are difficult to observe for potential employers. Data Appendix A.1 provides the definitions of these cognitive ability measures.

Non-cognitive traits include self-esteem, intrinsic motivation, extrinsic motivation, and the Big Five personality test (extraversion, openness, conscientiousness, agreeableness, and neuroticism). The additional baseline survey conducted in April–June 2015 collected data on risk and time preferences and rational decision-making ability using the tests recently developed by

[21] MSCE is an official test that all Malawian students must take to graduate from secondary school. AFF had access to the administrative MSCE score data via the cooperation of the Ministry of Education of the Republic of Malawi. We use Math and English test scores only because they are mandatory subjects of the MSCE test.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Choi et al (2014).[22]

Training outcomes are measured by a quiz score and the proportion of erroneous entries in a practice survey.[23] The quiz tested specific knowledge on the census details. It consists of 12 questions, a mixture of open-ended and true/false type questions. The full text of the quiz is presented in Figure A.4.

Main job performance measures during the census are survey quantity and quality. Survey quantity is measured by the number of households surveyed by each enumerator per day. Survey quality is measured by the proportion of systematically inconsistent or incorrect entries in the

[22] As explained in Subsection 2.2.1, risk and time preferences, and rational decision-making ability were measured after the census was completed. We included these measures in the randomization balance test under the assumption that these traits were not affected by our experiment. Data Appendix A.1 provides the details of how we measure them.

[23] The purpose of the practice survey was to practice interview skills before enumerators were dispatched to the field. The practice survey performance was evaluated as follows: First, we randomly matched two trainees. Each trainee in a randomly assigned pair received a pre-filled census questionnaire sheet and a blank survey questionnaire sheet. Then, one trainee interviewed the other matched trainee in the same pair and the latter trainee responded based on the assigned survey sheet. There were two different types of pre-filled questionnaire sheets with different hypothetical household information. Thus, trainees in the same pair acted as if they were two different households. Each trainee in every pair conducted this practice survey by changing roles. After conducting practice survey sessions, supervisors collected the survey sheets and calculated the error rate.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

census questionnaire specific to each household surveyed. For example, if a respondent has a child, the information about her child should be filled in. If not, it is counted as an error. Data Appendix A.2 provides the details about how we calculate the survey error rate. We also use subjective performance evaluations (SPE) measured by census respondents because we expect enumerators to give good impressions to community members as an NGO worker that serves local communities. During the PES, census respondents were asked to evaluate how carefully the enumerator had explained the questions.[24] In addition, after the completion of the census, 12 supervisors jointly evaluated the work attitude of each enumerator (SPEs measured by AFF supervisors).[25]

Lastly, census data were used to calculate the average characteristics of the catchment area so that we can use them as the control vector in the main regression analysis.[26]

---

[24] The question asked was "Whenever you were confused or could not understand the meaning of any question, did the enumerator carefully explain the meaning of the questions to you?". We analyze SPE by census respondents only when the census respondent and the PES respondent were identical. The probabilities that an original census respondent was a PES respondent are 77%, 77%, 83%, and 82% for *G1*, *G2*, *G3*, and *G4*, respectively. These rates are significantly different. Hence, the interpretation of the SPE analysis by respondents should be taken with caution.

[25] We asked a group of supervisors to evaluate general work attitude of enumerators. Enumerators were scored on a scale of 1 to 3.

[26] Regarding catchment area size, we could not acquire information on the exact land size of each catchment area. However, we had an unofficial, categorical measure of land size ranging from 1 (smallest) to 10 (largest), jointly determined by AFF supervisors who have worked in the Chimutu district for five years or longer.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Columns (2) and (3) of Table A.2 present the baseline characteristics of the *Internship* and *Wage* groups, respectively. The results of the first- and second-stage randomization balance are presented in Columns (4), (5) and (6). Panel A represents individual baseline characteristics of study participants. Study participants are about 20 years old and only 9% work in the official sector reflecting weak labor demand in Malawi.[27] Data Appendix A.1 provides the specific definition of the variables presented in Panel A. Panel B represents the catchment area characteristics where enumerators were dispatched. The results confirm that the study groups are well balanced: the proportion of statistically significant mean difference at the 10% significance level is 2 out of 28 (7.1%) in Column (4), 3 out 28 (10.7%) in Column (5), and 4 out of 28 (14.3%) in Column (6).

We also examine whether the baseline survey participants and non-participants are systematically different. Table A.3 shows that they are not statistically different from each other in most dimensions except for the household asset score. In addition, Table A.4 shows no systematic differences across enumerators assigned to each supervisor team, which confirms that the supervisor team randomization went well.

4. Main Results

4.1. Job Offer Take-up

Column (1) of Table 2 shows that the job offer take-up rates between the *Internship* and *Wage* groups are not statistically different. We test multidimensional sorting discussed in Dohmen

[27] The employment rate of baseline survey non-participants is similar. We reached non-participants via phone calls and 9.7% of them told us that they did not attend because they were working.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

and Falk (2011) by exploring whether career and wage incentives attract those with different observable characteristics. Columns (2) to (18) of Table 2 show the regression results of the following equation:

$$Accept_i = \alpha + \delta \cdot Internship_i + \lambda \cdot Trait_i + \varphi \cdot Internship_i \cdot Trait_i + \epsilon_i \quad (1)$$

where $Accept_i$ is a binary indicator that equals 1 if individual i accepted a job offer, and 0 otherwise. $Internship_i$ is a binary indicator if individual *i* belongs to the *Internship* group and the omitted category is the *Wage* group. $Trait_i$ is an individual characteristic variable that we evaluate one by one. ϵ_i is an error term. We test whether career incentives attract workers differently over a variety of individual characteristics including demographic and socioeconomic characteristics, cognitive ability index, and non-cognitive traits.

Our coefficient of interest is φ, which captures differential take-up of a job offer between the *Internship* group and the *Wage* group by individual traits. We find that none of the estimates of φ across individual traits is statistically significant at the 5% level.[28] These findings imply that observable characteristics are not likely to predict self-selection.

Table A.5 provides additional evidence on self-selection by comparing the observable

[28] There might be concern about statistical power due to relatively small sample size (N=362). However, for most variables we are able to detect 15% differences between the two groups. For example, Column 2 of Table 2 shows we are able to detect age difference between the two groups that is bigger than 0.07 (=0.037*1.96) years, which is a 0.36% change (=(0.07/20.4)*100). Nonetheless, we cannot fully rule out the possibility that we are unable to detect small differences between the two groups. Therefore, the results should be interpreted with this caveat.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

characteristics of job offer takers between the *Internship* group and the *Wage* group. The results in Table A.5 confirm the results in Table 2 that the two groups are not systematically different in terms of both statistical and economic significance.[29]

The absence of systematic differences in observable characteristics does not necessarily mean that unobservable characteristics, training outcomes, and job performance would be the same if some of the unobservable characteristics were to affect training outcomes and job performance.

4.2. Training Outcomes

Even though we do not find any differences in observable characteristics between job takers of the two groups, we might find a difference in training outcomes if career and wage incentives attract people with different unobservable characteristics. Panel A of Figure A.5 displays the kernel density estimates of the training outcomes measured by the quiz score and the practice survey error rate. Table 3 shows the corresponding results from the following specification:

$$Training_i = \alpha + \beta \cdot Internship_i + \omega_i \quad (2)$$

where $Training_i$ is the training outcomes such as practice survey error rate and quiz score for individual i. For the practice survey error rate regression, we control for a practice survey type and

[29] We acknowledge that study participants could have responded to the self-reported non-cognitive tests in a way that they believed to be desirable from the perspective of a potential employer, even though they were not aware of the possibility of a job offer at the time of the baseline survey. This is consistent with the real world in which job seekers are not able to manipulate test scores (cognitive ability) in a pre-employment test but might try to respond to a personality test in a way in which they have a desirable non-cognitive skill.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

pair–fixed effect in the regression.[30]

Panel A of Figure A.5 shows that the *Wage* group performs better than the *Internship* group in terms of both quiz score and practice survey error rate. Panel A of Table 3 provides corresponding results from the regression. It confirms that the quiz score of the *Internship* group trainees is 2.0 points (23.8%) lower than that of the *Wage* group trainees as shown in Column (1). Similarly, the survey error rate is 10.4 percentage points (38.2%) higher among the *Internship* group trainees than that among the *Wage* group trainees as shown in Column (3).

At the end of the training, AFF disqualified 11 trainees who did not meet the minimum qualification requirement. As the abovementioned regression results indicate, the *Internship* group performed worse than the *Wage* group did. Thus, all dropouts (11 trainees) came from the *Internship* group only. Panel B of Table 3 presents the training outcomes of enumerators dispatched to the field by excluding the 11 training failures. The regression results between the two panels are qualitatively similar, but the magnitude of the coefficient estimates is larger in Panel A than in Panel B, because those who failed training are all from the *Internship* group.

The specification used in Columns (2) and (5) is to test whether individual observable characteristics can explain the differences in the training outcomes between the two groups. The individual observable characteristics include age, household asset score, cognitive ability index, and non-cognitive traits, such as self-esteem, intrinsic and extrinsic motivation, and Big 5 personality scales. We find similar coefficient estimates between Columns (1) and (2). For

[30] All regressions include number of siblings, which is not balanced in the baseline, and eligibility for AFF's past interventions as a control vector. When analyzing the practice survey error rate, we additionally include survey pair fixed effect.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

example, observable characteristics explain only 2.5% (=(2.01-1.96)/2.01) of the difference in quiz score. In the case of the practice survey error rate, controlling for individual characteristics in Column (5) makes coefficient estimates statistically insignificant and larger. These findings imply that observable characteristics are somewhat limited in explaining the difference in the training outcomes.

In summary, we find that those attracted by a job offer with wage incentives outperformed those attracted by a job offer with career incentives in the training. This difference could be caused by workers with different characteristics selecting into different work incentives, thereby creating the difference in the training outcomes (selection effect).

However, there are several reasons why the observed difference in training performance could be different from the true selection effect. For instance, those in the Internship group have an incentive to exert more effort than the Wage group due to the future job prospect of the career incentives. That is, in the absence of such an effect, the difference in training performance due to selection could be larger than the observed difference in training performance. On the other hand, the difference in training performance due to selection could be smaller if there was a learning-by-doing effect for training instructors. Instructors could deliver lectures more efficiently in the second session (for the Wage group) than in the first session (for the Internship group). Therefore, the analysis of the training results should be interpreted with caution due to these possibilities that can potentially bias the selection effect

4.3. Selection effect of career incentives on labor productivity

In this subsection, we examine the selection effect of career incentives evaluated against wage incentives on job performance. As previously discussed, *G2* and *G3* have the same incentives at work, but the channels by which they were recruited are different. Therefore, we interpret that

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

differences in performance are driven by the selection effect.

Our identifying assumption is that G2 and G3 enumerators perceive their work incentives identical at work even though the sequences by which career and wage incentives were presented are different. The different sequence could form different perceived valuation of the incentives that affect enumerators' feelings leading to different levels of work efforts. As a result, our estimates of the selection effect would be biased as Abeler et al. (2011) discussed. However, we argue this is unlikely. If there were such a difference in feelings, we expect that differences in job performance would become smaller over time because the difference in feelings might diminish with time. Figure A.6 shows that the difference in job performance is fairly constant over time.[31]

Panel B of Figure A.5 suggests that *G2* has higher labor productivity than *G3* in terms of survey quality and quantity. This finding is surprising because the *Wage* group had better training outcomes than the *Internship* group did. We test this graphical evidence formally by estimating the following equation:

$$Y_{ijklt} = \alpha + \beta \cdot G2_j + \gamma \cdot H_{ik} + \varphi \cdot Z_k + V_{lt} + \sigma_t + \psi_{ijklt} \tag{3}$$

where Y_{ijklt} is job performance measured in the survey collected from household *i* by enumerator *j* whose supervisor is *l*, in catchment area *k*, surveyed on the *t*-th work day. $G2_j$ is 1 if enumerator *j* belongs to *G2* and 0 if he belongs to *G3*. H_{ik} is a vector of respondents' household characteristics

[31] The different sequence could still generate bias if those recruited with career incentives might misunderstand the addition of wage incentives as a reward for good performance during training, while those recruited with wage incentives might misunderstand the addition of career incentives as a windfall gain, not a reward. However, this is also unlikely because we clearly indicated that the additional provision of incentives in the second stage was randomly determined.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

and Z_k is a vector of catchment area characteristics.[32] V_{lt} is the supervisor team-specific post-visit effect and σ_t is the survey date fixed effect.[33] Standard errors are clustered at the catchment area level. For dependent variables, we use survey quality measured by the survey error rate ($Error_{ijktl}$) and survey quantity measured by the number of surveys per day ($Survey_{jktl}$).

Panel A of Table 4 presents the regression results from equation (3). We find that *G2* outperforms *G3* in two main measures of job performance, even though *G3* outperforms *G2* during the training. The error rate is 2.2 percentage points (28.6%) lower in *G2* than *G3*, as shown in Column (1). The survey quantity of *G2* is higher than that of *G3* by 1.39 households per day (13.0%), as shown in Column (4).

To assess how much observable individual characteristics and training performance can explain the selection effect estimated in Columns (1) and (4), we additionally control for enumerator characteristics such as demographic and socioeconomic status, cognitive ability (MSCE scores and Raven's matrices/O*NET scores), and non-cognitive traits in Columns (2) and (5) as well as training performance in Columns (3) and (6). As shown in Columns (2) and (5), observable individual characteristics of enumerators are limited in explaining the estimated selection effect. On survey quality, the inclusion of observed individual characteristics does not explain the estimated selection effect of career incentives at all. It explains survey quantity only

[32] Respondent's household characteristics include the fixed effect for family size. Catchment area characteristics include the total number of households, size of the catchment area, asset score, birth rate, malaria incidence, rate of birth with the assistance of a health professional, and death rate.

[33] $V_{lt} = \eta_0 + \eta_{1l} I(t > First) + \eta_{2l} I(t > Second)$ where *First* and *Second* are the dates of supervisor team *l*'s first and second visits, respectively, to enumerator *j*.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

by 7.2% (=(1.39-1.29)/1.39). Additionally controlling for training performance also remains limited in explaining the selection effects.

We present the selection effect on SPEs in Table A.6. *G2* has a 67.9% higher SPE score by survey respondents than *G3*, as shown in Column (1). Adding enumerator characteristics explains only 7.0% of the selection effect on SPE by respondents. This result is consistent with the fact that the observable characteristics of job takers between the *Internship* group and the *Wage* group are not different. Lastly, we find that the SPE score by supervisors is higher in *G3* than in *G2* (Column (4)), but it is not statistically significant at the 5% level. We do not control for σ_t and V_{lt} when we analyze SPE score by supervisors because it does not vary over time and catchment area.

In Table A.7, we report the results that decompose the main outcomes. To understand where survey errors come from, we decompose errors into incorrectly entered entries (e.g., filling in 179 for a person's age) and incorrectly missing entries (e.g., a child is present in the household but his/her age is missing). To better understand how survey quantity changes, we conduct regression analyses on three time-use variables such as total work hours per day, average survey time per household, and intermission time between surveys.[34] Column (3) in Panel A indicates that

[34] Work hours per day are the difference between the beginning time of the first survey and the end time of the last survey of the day. Intermission time is defined as the difference between the beginning time of a survey and the end time of the previous survey. The survey beginning and end times were recorded as a part of the census questionnaire. However, there was a sizable number of missing values, so we imputed those missing values (See Data Appendix A.3). The results remain similar even if we do not use the observations with imputed time values.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

the selection effect of career incentives on survey quality reported in Table 4 is mostly driven by the decrease in incorrectly missing entries. In addition, we find that the selection effect of career incentives on survey quantity comes from longer work hours, shorter survey time per household, and shorter intermission time as shown in Columns (5)-(10) of Table A.7. However, these coefficient estimates are not precisely estimated. We find that observable enumerator characteristics and training performance do not explain differences between *G2* and *G3* much.

Then, why do *G2* enumerators outperform *G3* enumerators in actual job performance, while the *Wage* group outperforms the *Internship* group during training? One possible explanation is that different skill sets are required in each setting. The test taken during the training was in a classroom setting, while job performance resulted from actual interactions with respondents in the field. It is plausible that enumerators selected through career incentives have comparative advantages in on-the-job performance but not in tests in a classroom setting. A critical characteristic of an enumerator is the skill to ask strangers sensitive questions about their households. This kind of skill might not be captured easily in a test taken in a laboratory setting.[35,36]

[35] Alternatively, it is possible that the *Internship* group initially had lower performance in the training but caught up with the *Wage* group later in the field owing to a steeper learning curve. However, this is less likely, as we find no evidence of performance catch-up. Job performance between the *Internship* and *Wage* groups remained constant over the study period (see Figure A.6 for the daily performance trend). It is also possible that screening out 11 trainees in the *Internship* group served as a reminder or a credible threat to those with career incentives that only some of them would be hired as regular workers in AFF, causing *G2* to work harder than *G3*.

[36] All 11 trainees who were dropped were from the *Internship* group. Therefore, if the labor

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

4.4. Incentive effects of work incentives on labor productivity

To measure causal impacts of career incentives on labor productivity, we compare job performance of enumerators who receive both wage and career incentives (*G3*) and that of enumerators with wage incentives only (*G4*). Similarly, we measure causal impacts of wage incentives by comparing job performance between enumerators with only career incentives (*G1*) and enumerators with both career and wage incentives (*G2*). We estimate incentive effects of wage and career incentives among job takers of the *Internship* and *Wage* groups, respectively; therefore, these incentive effects are not directly comparable. Panels B and C of Table 4 report the incentive effects of career and wage incentives on job performance estimated among the *Wage* and *Internship* groups, respectively. Panels C and D in Figure A.5 present the corresponding graphical evidence.

Our conceptual framework predicts that the additional provision of career incentives would motivate enumerators to exert more effort and improve job performance. However, in Panel B of Table 4, we find no such evidence in main labor productivity outcomes. However, column (4) of

---

productivity of the dropouts were lower than that of the hired enumerators, the performance-improving selection effects would be overestimated. However, we do not consider that any particular adjustment is necessary in the main analysis because screening out trainees who did not meet the minimum requirement is a regular business practice. Nevertheless, we re-estimate equation (3) after dropping 11 trainees with the lowest training scores from the *Wage* group (six from *G3* and five from *G4*). Panel A of Table A.8 shows that the results for the selection effects remain mostly robust; the size of the coefficients for the selection effect on survey quality becomes smaller, while that for survey quantity becomes larger. We find similar results on incentive effects (Panels B and C) and combined effects (Panel D).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Table A.6 shows that SPE measured by supervisors significantly increases by 51.5%. In summary, career incentives given to existing workers hired through the wage incentive channel do not improve labor productivity, but they induce enumerators to have better evaluation from supervisors. We speculate that the effort level of the *Wage* group enumerators was already high, and thus it is difficult for them to improve work performance at least in the short run. They rather exerted effort in building their relationships with supervisors.[37]

There might be a concern that, despite high frequency data, the relatively small number of enumerators allows for the detection of only relatively large effects and makes it difficult to interpret null results. Indeed, we are somewhat underpowered in the regression analysis of Panel B of Table 4 in the sense that the size of the standard errors is not small enough to capture the small effect (if any) of the work incentives. To illustrate, we are able to capture the impacts of career incentive on survey quality and quantity only if the change is greater than 16.7%

[37] Another possibility is that career incentives might not be very appealing to enumerators recruited through wage incentives conditional on self-selection. For example, enumerators might not have needed a job for a longer period. Alternatively, the marginal effects of career incentives in the second stage could be small, because enumerators had already received wage incentives in the first stage. However, this possibility does not explain an increase in SPE by supervisors. Lastly, there exists concern that the differences in performance could be driven by the decrease in control group productivity due to disappointment at not receiving the second-stage incentives. However, this possibility is less likely because this psychological mechanism, if present, would decline over time as such feeling might diminish with time, which does not correspond to the results shown in Figure A.6.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

(=0.007×1.96/0.082) and 13.0% (=0.763×1.96/11.5), respectively.

Panel C of Table 4 shows that wage incentives, additionally given to the *Internship* group enumerators, improve job performance. We find that survey errors decrease by 3.8 percentage points (a 50.1% decrease) in Column (1) without statistically significant changes in survey quantity (Column (5)) and SPEs (Panel C of Table A.6). Panel C of Table A.7 shows that the decrease in the survey error rate is explained mostly by a decrease in illogical missing entries, as shown in Column (3).[38] This finding is consistent with the gift exchange model of the efficiency wage theory formulated by Akerlof (1984). In the model, a worker exerts more efforts upon receiving a gift from an employer that exceeds the minimum level of compensation for the minimum level of effort. We also acknowledge that a part of the productivity improvements in *G2* (evaluated against *G1*) might not be completely due to the gift exchange motive because the wage incentives include a performance bonus component.

Panel D of Table 4, which compares *G1* versus *G4*, resembles the combined effects of selection and incentive effects on productivity in that participants were attracted to accept a job offer via different incentives and the incentives at work also remained different. It is noteworthy that the combined effects of career incentives (Panel D) are not necessarily a simple sum of the selection effect (Panel A) and incentive effect (Panel B), because of potential interaction between

[38] One might wonder that the *G1* enumerators who have career incentives only performed poorly due to lack of money for meals in the field. To minimize this possibility, AFF informed all enumerators in advance that it would be difficult to find a shop or restaurant in the field, and encouraged them to bring enough of their own food during the work period. AFF ensured that the enumerators were able to use the kitchen for cooking at the pre-arranged housing during the census.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

selection and incentive effects. In addition, the study sample used in Panel D of Table 4 is different from that in Panels A and B. We find no significant difference in the combined effects between *G1* and *G4* in the main productivity outcomes, implying the importance of separating selection and incentive effects. However, we find that *G1* enumerators have significantly better SPE by supervisors than *G4* enumerators do (Panel D of Table A.6), which is consistent with the fact that career incentives causally improve SPE by supervisors in Panel B.

5. Conclusion

This study analyzes how career and wage incentives affect labor productivity through a two-stage randomized controlled trial in the context of a recruitment drive for census enumerators in Malawi. Even though career and wage incentives are the most common types of work incentives, no study has considered these incentives in the same setting, to the best of our knowledge.

We find that career incentives of an internship significantly improve labor productivity through the self-selection of workers: The *Internship* group (those attracted by career incentives) outperformed the *Wage* group (those attracted by wage incentives) at work, even though the *Wage* group was better than the *Internship* group during the training. Observable individual characteristics, including training outcomes, are limited in explaining the difference in labor productivity. The fact that neither observable characteristics nor training outcomes predict actual job performance implies that screening via observable characteristics is imperfect, particularly when hiring entry-level workers who have no track record of past job history or credentials to verify their unobserved productivity. Furthermore, these findings highlight the importance of a recruitment strategy in attracting workers with strong unobservable skills via self-selection (e.g., an internship).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Regarding the career incentive effect, we find no positive evidence for the career incentive effects on labor productivity conditional on selection except for the SPE by supervisors. Our findings suggest that career incentives are effective in improving labor productivity mainly through the selection effect channel. Lastly, we find that additional financial incentives can be an effective means to improve labor productivity (e.g., survey quality) for those recruited by career incentives. As a result, labor productivity is highest in *G2*, who were recruited by career incentives and received additional wage incentive.

We show how work incentives affect labor productivity among entry-level workers in Malawi. Therefore, our setting is closest to situations in which firms hire entry-level workers in developing countries whose productivity is not easily observable and worker characteristics are similar due to the similarity in contexts. Our analysis has implications for settings in which employers have difficulties screening productive workers with no or short employment history and are looking for effective means to motivate existing workers.

There are limitations to our study. First, we acknowledge that the approach by which we estimate the incentive effects might not perfectly characterize the real world. In the real world, workers might not always receive additional incentives without prior notice. Second, the length of the job we study is relatively short-term. As such, we cannot study whether the estimated selection and incentive effects of career and wage incentives remain constant over longer periods. The short-term nature of our study also limits the analysis of the effects of work incentives on retention. Third, we do not directly observe the individual's perception of the value of work incentives. In addition, we do not measure how career and wage incentives change workers' belief about the probability of retention by AFF. Hence, we do not know whether the selection effect of career incentives operates through the expectation of a job prospect at AFF or a potentially favorable

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

recommendation letter. Fourth, the non-cognitive traits used in this study are self-reported psychometric scales measured based on a paper test. It would be interesting to know whether such paper-based and self-reported non-cognitive traits are highly correlated with non-cognitive traits measured in other settings. Fifth, the relatively small number of enumerators may prevent us from interpreting relatively small and insignificant effects, especially in estimating the career incentive effects. However, most major outcomes (selection effects and wage incentives effects) are large enough to detect their effects.

The difficulty in effective screening of job applicants and lack of motivation among existing workers are key drivers of low labor productivity, particularly in developing countries. A better understanding of selection and incentive effects of work incentives would allow employers to design optimal employment strategies. Based on our findings, we argue that active adoption of career incentives in the workplace as a hiring strategy could be an effective means to increase labor productivity of an organization hiring entry-level workers.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman, 2011. Reference Points and Effort Provision. *American Economic Review*, 101(2): 470–492.

Akerlof, George A., 1984. Gift Exchange and Efficiency–Wage Theory: Four Views. *American Economic Review: Papers and Proceedings*, 74(2): 79–83.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack, 2014. No Margin, No Mission? A Field Experiment on Incentives for Public Services Delivery. *Journal of Public Economics*, 120 (Dec 2014): 1-17.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee, 2016. Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery. Harvard Business School Working Paper.

Brooks, L., A. Cornelius, E. Greenfield, and R. Joseph, 1995. The Relation of Career-related Work or Internship Experiences to the Career Development of College Seniors. *Journal of Vocational Behavior*, 46(3): 332–349.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman, 2014. Who is (More) Rational? *American Economic Review* 104(6): 1518–1550.

D'Abate, C.P., M.A. Youndt, and K.E. Wenzel, 2009. Making the Most of an Internship: An Empirical Study of Internship Satisfaction. *Academy of Management Learning & Education*, 8(4): 527–539.

Dal Bó, Ernesto, Frederico Finan, and Martin A. Rossi, 2013. Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *Quarterly Journal of Economics*, 128(3): 1169–1218.

Deserranno, Erika, 2018. Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda. *American Economic Journal: Applied Economics*, 11(1): 277-317.

Dohmen, T. and A. Falk, 2011. Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*, 101(2): 556–590.

Duflo, E., R. Hanna, and S.P. Ryan, 2012. Incentives Work: Getting Teachers to Come to School.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

American Economic Review, 102(4): 1241–1278.

Fryer, Roland, 2013. Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*. 31(2): 373–427.

Gagliarducci, Stefano and Tommaso Nannicini, 2013. Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection. *Journal of the European Economic Association*.

Glewwe, P., N. Ilias, and M. Kremer, 2010. Teacher Incentives. *American Economic Journal: Applied Economics*, 2(3): 205–227.

Gneezy, Uri, and John A. List, 2006. Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments, *Econometrica*, 74(5): 1365–1384.

Guiteras, Raymond P. and B. Kelsey Jack, 2018. Productivity in Piece-Rate Labor Markets: Evidence from Rural Malawi. *Journal of Development Economics,* 131: 42–61.

Lazear, Edward P., 2000. Performance Pay and Productivity. *American Economic Review*, 90(5): 1346–1361.

Liu, Y., G.R. Ferris, J. Xu, B.A. Weitz, and P.L. Perrewé, 2014. When Ingratiation Backfires: The Role of Political Skill in the Ingratiation–Internship Performance Relationship. *Academy of Management Learning & Education*, 13(4): 569–586.

National Statistical Office (NSO), 2014. Malawi Labour Force Survey 2013. Zomba, Malawi.

Nunley, J.M., Pugh, A., Romero, N. and Seals Jr, R.A., 2016. College major, internship experience, and employment opportunities: Estimates from a résumé audit. Labour Economics, 38, pp.37-46.

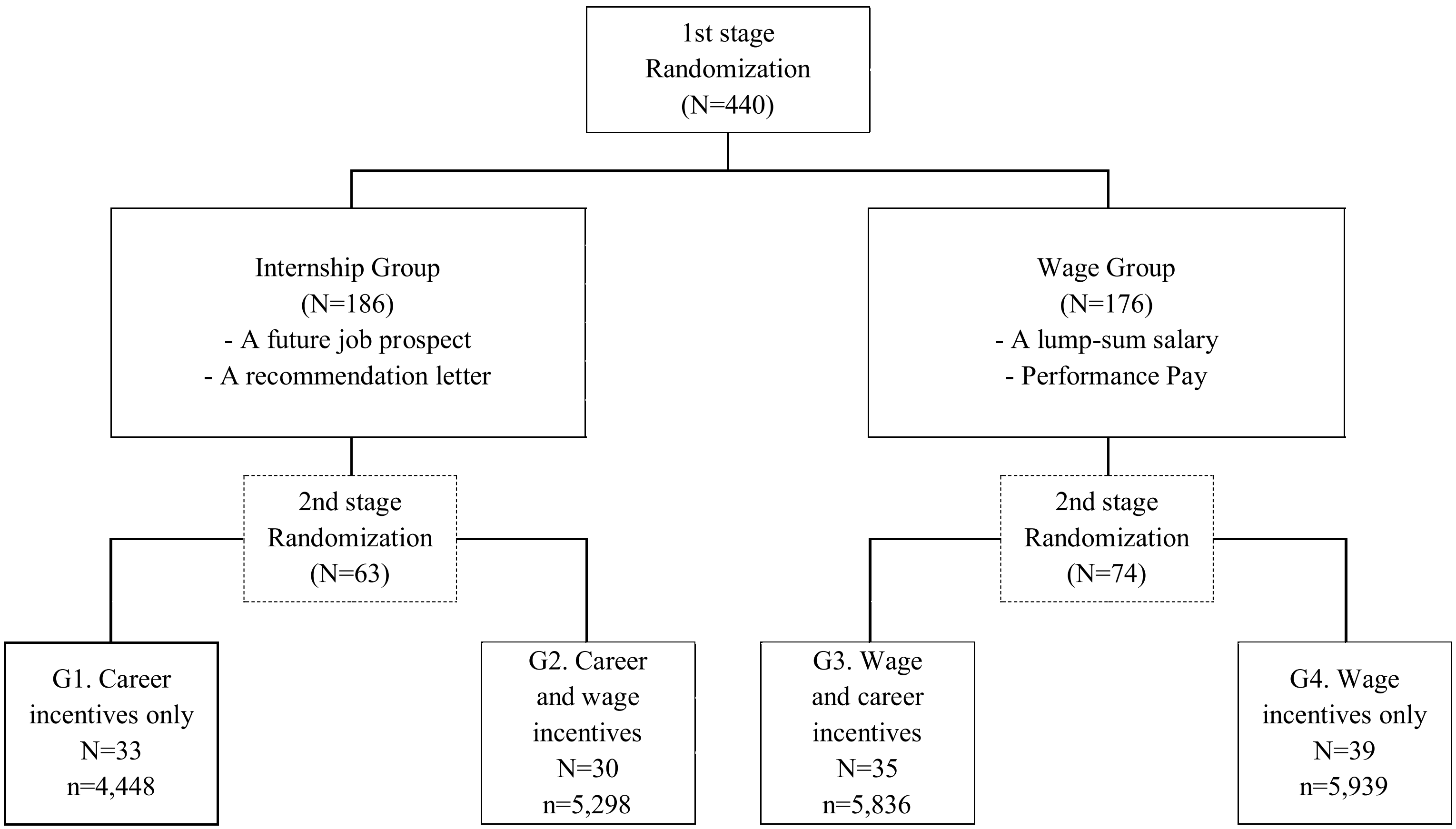
Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Shearer, B., 2004. Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *Review of Economic Studies*, 71(2): 513–534.

World Bank, 2016. World Development Indicators. http://data.worldbank.org/country/malawi

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Figure 1: Experimental Design



Notes: Upper case N indicates the number of participants in each stage. Lower case n indicates the number of surveys conducted by census enumerators

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Table 1 Experiment Stages

Stage of experiment			Number of individuals					
			G1 (career incentives only)	G2 (career incentives and additional wage incentives)	G3 (wage incentives and additional career incentives)	G4 (wage incentives only)	p-value	Total
A	Target study subjects	2011 Dec	220		220		-	440
B	Study participants (baseline survey participants)	2014 Dec	186 (84.1%)		176 (80.0%)		.265	362
C	Trainees	2015 Jan	74 (39.8%)		74 (42.0%)		.663	148
D	Trainees who failed training		11		0		-	11
E	Enumerators	2015	63 (33.9%)		74 (42.0%)		-	137

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

		Jan-Feb	33	30	35	39		
F	Number of surveys		4,448	5,298	5,836	5,939	-	21,521

Notes: The proportions of individuals remaining over experiment stages are in parentheses. The number of participants in the stage B is divided by the number of participants in the stage A, and the number of participants in the stages C and E are divided by the number of participants in the stage B.

Table 2 Job Offer Acceptance by Individual Trait

Dependent Variable (Job offer acceptance)	(1)	(2) Age	(3) Number of siblings	(4) Asset score	(5) Currently working	(6) Self-esteem	(7) Intrinsic motivation	(8) Extrinsic motivation	(9) Extroversion
Trait		.042	.038*	-.068*	-.107	-.024**	-.012	-.019	-.058*
		(.030)	(.019)	(.040)	(.136)	(.010)	(.108)	(.136)	(.032)
Internship group	-.024	-.323	-.029	-.023	-.025	-.321	.521	.733	-.297*
	(.052)	(.747)	(.131)	(.085)	(.055)	(.278)	(.491)	(.520)	(.173)
Trait *		.015	-.002	-.009	.028	.015	-.176	-.266	.077*

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Internship group		(.037)	(.028)	(.054)	(.180)	(.014)	(.157)	(.182)	(.046)
Constant	.481***	-.372	.326***	.558***	.491***	.931***	.517	.537	.683***
	(.055)	(.613)	(.094)	(.073)	(.057)	(.205)	(.336)	(.387)	(.126)
Observations	362	362	362	362	362	362	362	361	358
R-squared	.018	.046	.036	.036	.021	.034	.027	.031	.027
Mean (SD)		20.4(1.65)	4.39(1.80)	1.14(.896)	.086(.280)	19.3(3.69)	3.09(.340)	2.84(.282)	3.54(1.16)

Dependent Variable (Job offer acceptance)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Agreeableness	Conscientiousness	Emotional stability	Openness to experiences	Time preference	Risk preference	Rational decision-making ability	MSCE score	Raven and O*NET score
Trait	-.001	.046*	.011	-.001	.196	.288	-.019	-.051	-.140***
	(.027)	(.026)	(.027)	(.027)	(.284)	(.498)	(.274)	(.040)	(.053)
Internship group	.025	.251	.145	.041	-.096	.388	-.228	-.028	-.035
	(.196)	(.216)	(.195)	(.187)	(.158)	(.413)	(.305)	(.052)	(.052)

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Trait * Internship group	-.010	-.049	-.033	-.013	.199	-.644	.257	-.033	-.050
	(.037)	(.037)	(.037)	(.035)	(.384)	(.640)	(.363)	(.056)	(.071)
Constant	.486***	.223	.426***	.485***	.407***	.299	.502**	.483***	.496***
	(.148)	(.152)	(.148)	(.148)	(.130)	(.324)	(.234)	(.055)	(.053)
Observations	362	361	360	362	334	335	334	362	362
R-squared	.019	.026	.020	0.019	.024	.019	.019	0.033	.069
Mean (SD)	5.11(1.39)	5.68(1.35)	5.07(1.45)	5.36(1.35)	.396(.144)	.635(.083)	.826(.149)	-.013(.857)	.037(.658)

Notes: Robust standard errors are reported in parentheses. ***, **, and * denote the significance level at 1%, 5%, and 10%, respectively. Asset score is the sum of items owned out of improved toilet, refrigerator, and bicycle. See Data Appendix A.1 for the definitions of MSCE score, Raven and O*NET score, and non-cognitive trait variables.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Table 3: Training Performance

Dependent variable	Quiz score		Practice survey error rate		
	(1)	(2)	(3)	(4)	(5)
Panel A: 148 Trainee Sample					
Internship group	-2.01***	-1.96***	.104***	.089***	.323
	(.344)	(.303)	(.026)	(.029)	(.206)
Observations	148	148	148	148	148
R-squared	.228	.534	.114	.239	.811
Wage Group Mean (SD)	8.43 (1.82)		.272 (.142)		
Panel B: 137 Enumerator Sample					
Internship group	-1.44***	-1.47***	.094***	.080***	.302
	(.329)	(.286)	(.028)	(.030)	(.210)
Observations	137	137	137	137	137
R-squared	.163	.511	.099	.243	.862
Wage Group Mean (SD)	8.43 (1.82)		.272 (.142)		
Individual characteristics	No	YES	No	No	YES
Practice survey pair FE	No	No	No	YES	YES

Notes: Robust standard errors are reported in parentheses. ***, **, and * denote the significance level at 1%, 5%, and 10%, respectively. All specifications (columns 1-5) include the number of siblings and binary indicators for previous AFF programs. The practice survey error rate regression includes a binary indicator for the survey questionnaire type. Columns 2, 4, and 5 include age, asset score, MSCE score, Raven and O*NET score, and a set of non-cognitive traits

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

(self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items). Column 5 includes dummies for each trainee pair who conducted the practice survey with each other.

Table 4 Selection and Incentive Effects of Work Incentives on Job Performance

VARIABLES	Survey quality (error rate)			Survey quantity (number of surveys per day)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Selection effect (G2 vs G3)						
G2	-.022**	-.023**	-.023**	1.39**	1.29**	1.09*
	(.009)	(.009)	(.009)	(.610)	(.542)	(.611)
Observations	11,130	11,130	11,130	1,003	1,003	1,003
R-squared	.162	.307	.308	.145	.170	.180
Mean (SD) of G3		.077 (.078)			10.7 (5.45)	
Panel B: Incentive effect of career incentives (G3 vs. G4)						
G3	.007	.006	.006	-.763	-1.14*	-1.14*
	(.009)	(.010)	(.010)	(.681)	(.628)	(.613)
Observations	11,775	11,775	11,775	1,063	1,063	1,063
R-squared	.189	.269	.276	.152	.195	.199
Mean (SD) of G4		.082 (074)			11.5 (6.36)	
Panel C: Incentive effect of wage (G1 vs. G2)						
G2	-.038**	-.022**	-.019*	1.05	.644	.247
	(.016)	(.010)	(.010)	(.879)	(.941)	(.999)

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_00854

Observations	9,779	9,779	9,779	914	914	914
R-squared	.178	.357	.358	.203	.232	.242
Mean (SD) of G1		.075 (.068)			9.84 (5.19)	
Panel D: Combined effect (G1 vs. G4)						
G1	-.001	-.003	-.005	-1.41	-.732	-.259
	(.015)	(.013)	(.013)	(1.31)	(1.18)	(1.06)
Observations	10,424	10,424	10,424	974	974	974
R-squared	.194	.276	.277	.157	.232	.235
Mean (SD) of G4		.082 (074)			11.5 (6.36)	
Individual characteristics	NO	YES	YES	NO	YES	YES
Training performance	NO	NO	YES	NO	NO	YES

Notes: Robust standard errors clustered at the catchment area level are reported in parentheses. ***, **, and * denote the significance level at 1%, 5%, and 10%, respectively. All specifications (columns 1−6) include the number of siblings, catchment area characteristics, supervisor team-specific post-visit variables, survey date-fixed effect, and binary indicator variables for previous AFF programs. Catchment area characteristics include the total number of households, catchment area size, family size, asset score, number of births in the last 3 years, incidence of malaria among children under 3, and deaths in the last 12 months. Columns 2, 3, 5, and 6 include age, asset score, MSCE score, Raven and O*NET score, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items). Columns 3 and 6 additionally include the two measures of training performances: the quiz score and practice survey error rate.